AUGMENTED AUDIFICATION

Robert Höldrich, Katharina Vogt

Institute of Electronic Music and Acoustics,
University of Music and Performing Arts Graz, Austria
hoeldrich@iem.at, vogt@iem.at

ABSTRACT

We present a sonification method that blends between audification and auditory graphs which we call "Augmented Audification". It is based on a combination of single-side-band modulation and a pitch modulation of the original data stream. Benefits include the flexible adjustment of the sonification's frequency range to the human hearing range and the possibility to interactively zoom into the data set at any scale. The paper introduces the method by three examples: deterministic harmonic complexes, random signal analysis, and seismology.

1. INTRODUCTION

The choice of the sonification method for an existing problem is highly dependent on the size of the original data, as discussed within the concept of the sonic design space map (SDSM) by de Campo [1]. Taking on this perspective, we may place typical data sets of auditory graphs and ones of audification on opposite sides of the SDSM. Both methods belong to the standard repertoire of sonification design, and seem to exclude each other. Auditory graphs are usually very basic in their implementation, but work well for a wide range of data sets and users. Classically, they use a pitch-time-display for relatively few data points. Audification, on the other end of the scale, often runs with data display rates (sampling rates) of tens of kHz and needs accordingly large data sets.

This paper suggests a seamless interpolation between audification and auditory graphs. The method builds on well-known techniques of signal processing: single-side-band modulation utilizing the Hilbert transform, and frequency modulation. Basic features of the method are:

- fundamental properties of audification are conserved, notably the compact temporal support and the translation of high frequency content of the data into transient events in the sound
- independently to the rate of the data display, both the mean position of the frequency range *and* the bandwidth of the sonification can be controlled by free parameters
- data sets can be explored interactively at various time scales and in different frequency ranges.

In the following section, we will discuss basic properties and limits of audification and some aspects of auditory graphs. In Sec. 3, the signal processing algorithms of the proposed method are presented. Sec. 4 discusses three examples of the use of Augmented Audification. Finally, we give conclusions and an outlook to the further research agenda. Sound examples can be found at: http://iaem.at/Members/vogt/augmentedaudification

2. DISCUSSION OF EXISTING METHODS: AUDIFICATION AND AUDITORY GRAPHS

2.1. Audification

Audification is one of the oldest methods of sonification research. A prominent, early research example of audification has been reported by Speeth et al. [2] in 1961 who did a study on distinguishing audified seismic signals of earthquakes from ones of atomic bomb tests. Subjects showed high discrimination rates up to of 90%. The original definition of audification within a systematic sonification research was given by Kramer in 1991 (cited in [3], p. 186) and is still valid: "a direct translation of a data waveform to the audible domain".

A crucial advantage of audification is the following: By conserving the time regime of the data signal, audications of real physical processes are usually broad-band with a pronounced proportion of high frequencies during rapid transients. In the task of identifying natural sounds, e.g., the attack of musical instruments or speech signals, the transient signal portions provide important and salient features for the human ear and thus should serve as basis for pattern detection or recognition tasks in the auditory data exploration.

Many authors, e.g., Dombois and Eckel [4], have argued in favor of a puristic approach to audification with least data preprocessing as possible. This strategy should maximize the potential of the human hearing to detect yet unknown structures in the data which might be impaired by more sophisticated preprocessing. The only manipulation recommended within their narrow definition of audication is the variation of sampling frequency, i.e. the playback rate.

As long as the data array reflects the sampling of a band-limited physical process, a resulting sound signal corresponds to the data one-to-one. However, it should be noted that even a direct playback of the data may lead to a sound signal with information partly spoiled: an example are price trends of the financial market, where maximum and minimum prices serve as specific indicators. Because of the band-limited interpolation of the reconstruction filter of the D/A converter, the extreme data values would be super-elevated in the audification. In such a case, specific methods of interpolation as, e.g., Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) [5] have to ensure the conservation of the extrema.

The ideal audification signal has relevant auditory gestalts within time and frequency regimes that can be well-perceived by the human auditory system. This shall be shown with a numeric example: let us assume a low-pass data stream with transient events that appear within a range of 1k data points and with an (aperiodic) interval of roughly 10k data points. With a playback rate of 44.1kHz, we find approx. four of these events per second, which is comparable to the rate of syllables per second in English spoken language[1] and thus apt for human hearing. On the other hand, each transient event lasts for approximately 22ms and appears as a band-limited impulse with a cut-off frequency of around 50Hz, which is obviously far below the most sensitive frequency range of the hearing system. If the playback rate were to be raised by, e.g., a factor of 10 to 20, the individual impulses would be transposed to a more appropriate frequency range, but at the cost of an indiscernible temporal structure of the impulse series. Concluding from this example, pure audification suffers from a general trade-off between the macroscopic time scale and the frequency range of the relevant information.

Different concepts have been elaborated to cope with this trade-off. Worrall [7] extends the notion of audification, and allows other means of data pre-processing: besides filtering and data interpolation, i.e. compression and frequency shifting. This wider definition of audification still excludes the explicit synthesis of sound or the use of specific signal models. Worrall [7] also developed the Homomorphic Modulation Sonification, the mapping of sample values to pitch (see [8]).

2.2. Auditory Graphs

Just as audification, auditory graphs belong to the standard repertoire of sonification research since its beginning. Obvious benefits are the straightforward analogy to visual graphs, which make them intuitively understandable, at least for sighted users. Main application areas are accessibility and didactic examples, and they often serve as a basis for further, e.g., perceptual, sonification research. The data sets used are normally small, up to a few hundred data points.

The Sonification Sandbox [9] was possibly the largest effort to develop a general tool for auditory graphs. The tool allows to link any spreadsheet data to midi output and set various sound parameters (concerning timing, timbre, etc.). Development of the software is not maintained any more, making its use difficult for practical reasons. From the experience with the toolbox it can be concluded that most real-world sonification applications need a more flexible software environment. Still, the basic auditory graph has developed into a standard example of sonification.

Flowers [10] discussed promises and pitfalls of auditory graphs. He suggested the following strategies for successful displays:

- numeric values should be pitch-encoded
- the temporal resoluation of human audition shall be exploited
- loudness changes in a pitch mapped stream shall be manipulated in order to provide contextual cues and signal-critical events
- distinc timbres shall be chosen in order to minimize stream confusions and unwanted perceptual grouping
- time in sound shall be used to represent time in the data

All but the last two points are taken into account in the design of the proposed method: the last point, using time to represent time, might be fulfilled depending on the data set; the second last is only valid if several streams (of different data sets) are played in parallel, which is not intended for Augmented Audification.

3. AUGMENTED AUDIFICATION: THE MODEL

For explaining Augmented Audification (henceforth: AugAudif), we start with a basic audification. We assume a dataset $x(n)$ with $n = 1..N$ data points and a playback rate or sampling frequency f_p, i.e., f_p data points are displayed per second. The rendering over a D/A converter with a reconstruction filter leads to a continuous signal $x(t)$ with a Bandwidth B between zero and $1/2f_p$ Hz. If the playback rate is as low as a few hundred data points per second, the resulting sound will be in a low frequency range, where the human ear is not very sensitive.

3.1. Frequency Shifting

Therefore, as a first step, we perform frequency shifting using a single-side-band modulation. Using a Hilbert transform (see, e.g., [11]), the original audification signal $x(t)$ becomes the complex-valued signal $x_a(t)$:

$$x_a(t) = x(t) + j\ \mathcal{H}\{(x(t)\} \qquad (1)$$

with the imaginary constant j. This analytical signal can be written using a real-valued envelope $env(t) = |x_a(t)|$ modulated by a phasor with the instantaneous phase $\theta(t) = angle[x_a(t)]$:

$$x_a(t) = env(t)\ e^{(j\theta(t))} \qquad (2)$$

Performing a frequency shift by Δf and taking the real part of this signal leads to a SSB-modulated sound signal $x_{SSB}(t)$:

$$\begin{aligned} x_{SSB}(t) &= Re\left[env(t)\ e^{j(\theta(t)+2\pi\Delta f t)}\right] && (3)\\ &= x(t)\cos(2\pi\Delta f t) - \mathcal{H}\{x(t)\}\sin(2\pi\Delta f t) && (4)\end{aligned}$$

The spectrum of the analytical signal, which contains (only non-negative) frequencies between zero and B Hz, is shifted to the range between Δf and $(\Delta f + B)$). Discarding the imaginary part re-builds a symmetric spectrum.

The frequency shift Δf is a free parameter of the method, which helps to yield a perceptually optimal frequency range of the sonification, i.e., somewhere within the range of 100Hz and 2kHz.

If $\Delta f = 0$, there is no difference to a pure audification.

In the case of high playback rates, e.g., $f_p = 20$ kHz, which lead to a broad-banded audification, a frequency shift of $\Delta f = 100$ Hz hardly changes the overall signal, but might make low frequency components of the signal audible, as the spectrum is now shifted to the range between 100Hz and 10.1kHz.

A strong frequency shift, especially in combination with slow playback rates, results in a very narrow-banded signal which might be problematic from a perceptual point of view. The frequency shift squeezes the original - conceptually infinite - pitch range to a range of $(\Delta f + B)/\Delta f$. For example, if $f_p = 200$Hz, hence the bandwidth of the primary audification signal is max. 100 Hz, and the spectrum is shifted by $\Delta f = 500$Hz, the resulting bandwidth

[1] This is just a rule of thumb, not going into details of speech research, e.g., information aspects of tempo [6].

is 500 to 600Hz. Speaking in musical terms, all frequency components of the original data stream are now concentrated within a minor third. Fluctuations of such narrow-banded signals are difficult to perceive.

3.2. Exponential Frequency Modulation

Therefore the method is extended by modulating the frequency of the phasor of the analytic signal $x_a(t)$. The instantaneous frequency of the modulator, $f_i(t)$, exponentially encodes the numeric data values of $x(t)$ as pitch, following to Flowers' recommendations:

$$f_i(t) = 2^{cx(t)} f_0 \tag{5}$$

f_0 is the carrier frequency and c a freely choosable parameter that controls the magnitude of the modulation:

Setting $c = 0$ results in a constant instantaneous frequency of the frequency modulation which is then independent of the data values $x(t)$. This results in a pure frequency shift as described in Sec. 3.1.

Setting $c = 1$ leads to a transposition of one octave higher/lower for signal values $x(t) = +1/-1$.

For the AugAudif, the parameter of frequency shift is used as carrier frequency, $f_0 = \Delta f$. Integrating over the instantaneous frequency results in the instantaneous phase $\phi_i(t)$, which serves as a phase modulating term for the analytical signal.

$$\phi_i(t) = \int_0^t 2\pi\, \Delta f\, 2^{cx(\tau)} d\tau \tag{6}$$

The complete model of Augmented Audification is thus defined by:

$$\begin{aligned} x_{AA}(t) &= Re\left[env(t)\, e^{j\,(\theta(t)+\phi_i(t))}\right] && (7)\\ &= x(t)\cos(\phi_i(t)) - \mathcal{H}\{x(t)\}\sin(\phi_i(t)) && (8) \end{aligned}$$

The model is controlled by two freely choosable model parameters, Δf and c, that can be set according to the explorative goals of the sonification.

3.3. Implementation in SuperCollider

We implemented the AugAudif both using MatLab and SuperCollider, yielding the same results. MatLab allows for an analytic use of the method, thus we prepared the sound examples and plots discussed in the following in MatLab. SuperCollider (SC), on the other hand, is more handy for real-time, interactive use of the method.

We present the basic SC Code within this paper because of its simplicity, see Fig. 1. The implementation of AugAudif in SC starts from a given buffer buffer `b`, with an adjustable playback `rate` and a start position `startpos` from which the buffer readout starts. The model parameters are called `deltaf` and `c` according to the model definition. The instantaneous frequency `fMod` is defined, its sine and cosine calculated. The existing unit generator `HilbertFIR` returns a two-dimensional array: `hilb`[0] contains the primary signal `sig` (and is multiplied by the cosine), `hilb`[1] contains its Hilbert transform (which is multiplied by the sine). The final output is the difference between those two, according to Eq. 8.

```
SynthDef(\augaud,
    {|rate=0.05, c = 0.25, b=0, startpos=0, deltaf = 0|
    var fMod, sin, cos, hilb, sig;
    sig = PlayBuf.ar(1, b, rate:rate, startPos:startpos);
    fMod = deltaf * (2 ** ((sig * DC.ar(c))));
    sin = SinOsc.ar(fMod);
    cos = SinOsc.ar(fMod, 0.5pi);
    hilb = HilbertFIR.ar(sig, LocalBuf(4096)) * [cos, sin];
    Out.ar(0, hilb[0] - hilb[1]);
}).add;
```

Figure 1: Synth definition for a AugAudif implemented in SuperCollider.

In order to interactively test the parameter settings with different data sets, we created a simple GUI, see Fig. 2: It gives a visual overview of the data set, with a slider indicating the playback position, which also allows to change the starting position of the playback. The independently choosable parameters are the playback rate (by a factor relative to the sampling rate, usually 44100 Hz), the shifting parameter Δf determining the frequency shift of the Hilbert transform in Hz and the modulation factor c in octaves (and its equivalent in semitones). This simple GUI proved very efficient for testing the method with various data sets.

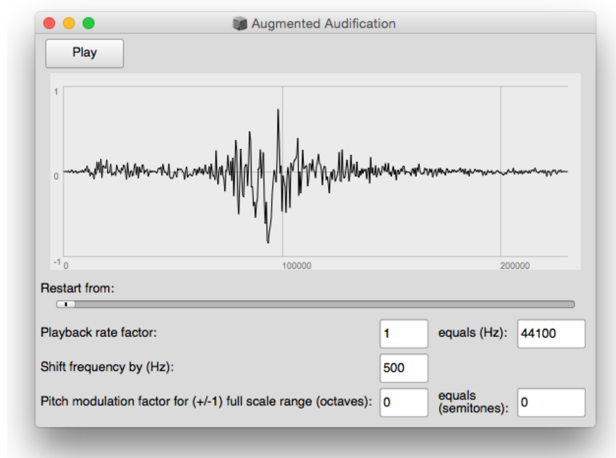


Figure 2: Graphical user interface within SuperCollider to quickly test different data sets and determine the ideal settings of model parameters.

4. MODEL FEATURES AND EXAMPLES

AugAudif allows to interpolate seamlessly between pure audification and an auditory graph in form of a pitch-time-display. As opposed to pure audification, where only the playback rate can be changed, two more model parameters can be chosen independently in the AugAudif. One parameter, Δf, controls the magnitude of a frequency shift. The second, c, sets the excursion of the exponential frequency modulation (FM), i.e. pitch modulation. For high playback rates, where pure audification might be used as well, the frequency shift has the advantage to bring low-frequency components of the spectrum within the human hearing range. For slow playback rates, the additional pitch modulation opens the possibility to perceptually zoom into the data due to the modulation-induced spectral spread.

The method preserves many advantages of audification. Large values of absolute magnitude of the data result in prominent epochs, thereby considering Flowers recommendation to "manipulating loudness changes in a pitch mapped stream to provide contextual cues and signal critical events". Furthermore, rapid transients in the data values lead to fast changes in instantaneous frequency and hence a short-time broad-banded sonification signal.

4.1. Example 1: Harmonic complex

Many physical processes are - at least approximately - periodic. The related signals therefore consist of harmonic partials, and their audification makes use of the human audition which groups these frequency components into one auditory *gestalt*. In pure audification, frequency ratios and thus the periodicity of harmonic complexes are preserved, resulting in one "sound" with a certain timbre and pitch. In AugAudif on the contrary, the frequency shift destroys the harmonic relationship between the partials and thus the periodicity of the signal. This results in a complex superposition of individual sinusoidal tracks instead of one gestalt with a certain timbre.

Let us consider a simple data signal, consisting of a cosine wave with constant frequency:

$$x(t) = cos(2\pi f_1 t) \tag{9}$$

The frequency shift with a certain Δf leads to another cosine wave (neglecting the modulation parameter, $c = 0$):

$$x_{AA}(t) = cos\left(2\pi\left(f_1 + \Delta f\right)t\right) \tag{10}$$

In this case, the augmentation leads to no further information. On the contrary, different values of f_1 might be harder to differentiate because they are shifted into higher frequencies. If $c > 0$, the pitch modulation produces side bands of the modulation, depending on the magnitude of f_1. If f_1 is small, up to about 15 Hz, a vibrato is perceived. Much higher frequencies produce typical FM-spectra as used in electronic sound synthesis [12]. Higher values of the modulation index c lead to pronounced side-bands and - depending on the specific values of f_1 and Δf - to inharmonic complex sounds that are perceptually not intuitive. Therefore c should be set carefully, especially for slow playback rates or if the audification signal is bandlimited for other reasons.

4.2. Example 2: Statistical properties of random data time series

Frauenberger et al. [13] reported a study on the audification of random data time series with varying higher order momentums. The third moment, skewness, is a measure for the asymmetry of the probability density function. The fourth moment is called kurtosis and serves as a measure of the peakedness of the distribution. In the study it has been shown that test participants could discriminate a kurtosis difference in the audification of above five. Qualitatively, the subjects reported an increase of roughness with rising kurtosis. Distinguishing different values of skewness could not be proven. This is not surprising, as skewness depends strongly on the mean of the data series which results in an indiscernible DC value, and furthermore, the hearing system does not perceive the sign of a signal.

We explored AugAudif with random data time series data in a preliminary test. A formal listening test is not within the scope of this paper, but discussed in the outlook. Subjective listening seems to show a much lower threshold for discriminating kurtosis and even the ability to defer different values of skewness using AugAudif. Two sound examples shall illustrate this finding:

Soundfile 1 is an AugAudif of time series with a white noise spectrum, zero skewness and varying kurtosis (consecutively 1, 2, 4, and 8). The playback rate of the data has been chosen as 800 Hz. Fig. 3 shows the spectrograms of these four sounds. (All spectrograms are logarithmic, between 0 and 10 kHz, calculated using a 4096 sample-Hanning window.)

Soundfile 2 is an AugAudif of time series with constant kurtosis but varying skewness. The parameters are the same as above ($f_p = 800$ Hz; $c = 5/12$; $\Delta f = 600$ hz). Kurtosis is set at 12, while skewness takes the values of -2, 0, and 2, respectively. The spectrograms shown in Fig. 4 clearly indicate the varying frequency excursions due to the different skewness.

A first, informal listening test of the authors of this paper showed clear differences in the resulting sounds of AugAudif of noises with kurtosis values above and below 3 (which corresponds to Gaussian noise and is often taken as a reference value in mathematics).

4.3. Example 3: Seismological data

As an example with real scientific data we take seismological data, stemming from Incorporated Research Institutions for Seismology [14].

The first example, *Seismo1* is an audification of a seismological event with 5s length given $f_p = 44100$Hz. Fig. 5(a) shows the spectrogram. (The waveform of this example is plotted within the GUI in Fig. 2.) The event is characterized by an impulsive sequence in the beginning with a bandwidth of around 5 kHz. The rest of the example shows relevant signal energies within 600 Hz bandwidth. The first half second and versus the end, we find high energies at very low frequencies that are hardly perceivable in the pure audification.

As a next step, the playback rate is reduced and the frequency is shifted in order to stay in the hearable range: the sound example is given for a deceleration factor of 4 ($f_p = 5.5$kHz) and $\Delta f =$ 250Hz. A little pitch modulation is added as well: $c = 1/12$; see Fig. 5(b).

Finally, Fig. 5(c) clearly shows the effect of the modulation. The modulation factor leads to a pitch transposition of a minor third for signal values of 1. When reaching the main impact of the event (possible the seismic shock), in this setting after around 12 seconds, the sound behaves clearly as an auditory graph. The glissando in the very beginning, which was out of the perceptional range in the pure audification, is now clearly audible.

The second example (from the same data source as the above, *Seismo2*) is 4s long using a playback rate of 44100 Hz. It contains a sequence of short "rattling" impulses (broad-band) embedded in a noisy background with energies predominantly in the frequency band between 100Hz and 2kHz. The spectrogram plots are given in Fig. 6. Starting from the final settings of the above example we slightly change the parameters to $f_p = 2.75$kHz, $\Delta f = 250$Hz, and $c = 1/12$ in order to make the transient events clearly audible. The modulation leads rather to noisy FM-synthesis sound (see Fig. 6(b) as compared to the pure audification in Fig. 6(a)).

The sound files and plots described above may give an idea of the potential of the proposed sonification method and the resulting sounds. Still, a main factor of the method is the interactive handling of the data set at various magnitudes.

5. CONCLUSION AND RESEARCH AGENDA

We presented Augmented Audification as a method that allows to blend between a pure audification and a pitch-based auditory graph. The implementation is simple and preserves preferable properties of audification. Furthermore, compared to pure audification the resulting sound may be better adapted to the human hearing range, since model parameters can be chosen independently from the sampling rate. Utilizing a frequency shift and an exponential frequency modulation, the data structures at various time scales can be made audible, permitting a true "zooming" and enabling the interactive exploration of a data set. Caution has to be taken with data sets that contain harmonic partials. While these partials would be perceived as a uniform auditory gestalt in pure audification, in our method, because of the frequency shift, the partials become perceptually separated.

This paper illustrates the main properties of the method through preliminary examples. Especially during the exploration of time series comprised of random data, the proposed method shows promising results as it succeeds in discriminating data series by their higher order statistical moments. Nevertheless, formal listening tests have to be performed in order to determine psychometric functions for the audibility of skewness and kurtosis, and to further elude the potentials and limitations of Augmented Audification in this field.

A second research thread would be defined by the exploration and further development of the method's interactivity. Zooming into data sets is difficult to realize in many sonification designs. This specific benefit of Augmented Audification should be studied using different data sets and applications. The pilot GUI as described in this paper will be refined and will be used to collect quantitative and qualitative data on sonification and user behavior.

6. ACKNOWLEDGMENT

We would like to thank Hanns Holger Rutz for feedback with the SC implementation.

7. REFERENCES

[1] A. de Campo, "Toward a data sonification design space map," in *Proceedings of the 13th International Conference on Auditory Display, Montreal*, 2007.

[2] S. Speeth, "Seismometer sounds," *J. Acous. Soc. Amer.*, vol. 33, pp. 909–916, 1961.

[3] G. Kramer, Ed., *Auditory Display. Sonification, Audification and Auditory Interfaces.*, ser. Proceedings Volume XVIII. Santa Fe Institute, Studies in the Sciences of Complexity, 1994.

[4] F. Dombois and G. Eckel, *The Sonification Handbook*. Logos Publishing House, Berlin, 2011, ch. 12 - Audification, pp. 237–272.

[5] F. N. Fritsch and R. E. Carlson, "Monotone piecewise cubic interpolation," *SIAM J. Numerical Analysis*, vol. 17, pp. 238–246, 1980.

[6] S. A. J. Wood, "Speech tempo," *Working papers, Phonetics Laboratory, Department of General Linguistics, Lund University*, 1973.

[7] D. Worrall, "Sonification and information – concepts, instruments and techniques," Ph.D. dissertation, University of Canberra, 2009.

[8] ——, "The use of sonic articulation in identifying correlation in capital market trading data," in *Proceedings of the 15th International Conference on Auditory Display (ICAD)*, 2009.

[9] B. K. Davison and B. N. Walker, "Sonification sandbox reconstruction: Software standard for auditory graphs," *Proc. of the International Conference on Auditory Display*, 2007.

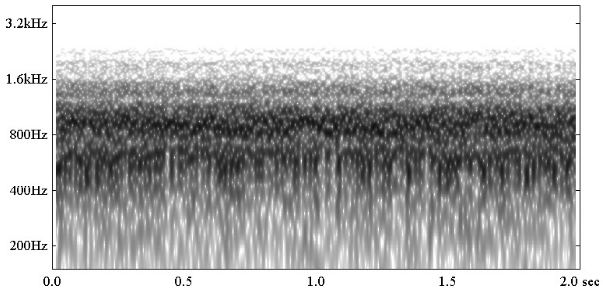
[10] J. H. Flowers, "Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions," *Proceedings of International Conference on Auditory Display*, 2005.

[11] A. Oppenheim and R. Schafer, *Digital Signal Processing*. Prentice Hall, 1975.

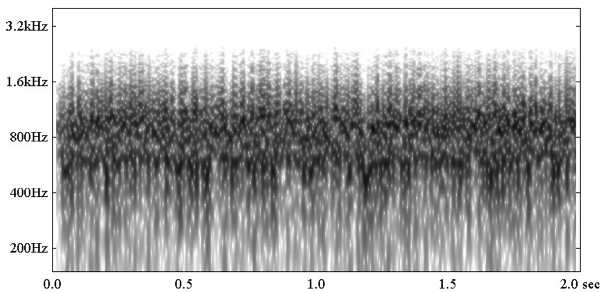
[12] C. Roads, *The Computer Music Tutorial*. MIT Press, Cambridge, Massachusetts, 1996.

[13] C. Frauenberger, A. deCampo, and G. Eckel, "Analysing time series data," in *Proc. of the 13th International Conference on Auditory Display*, 2007.

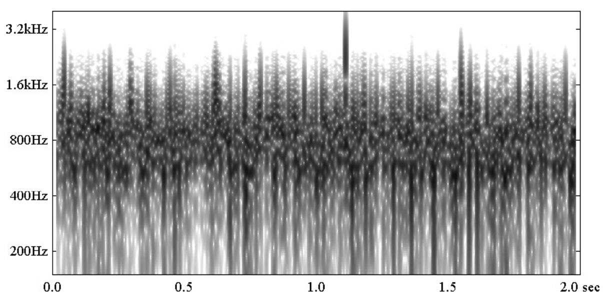
[14] [Online]. Available: http://ds.iris.edu/ds/



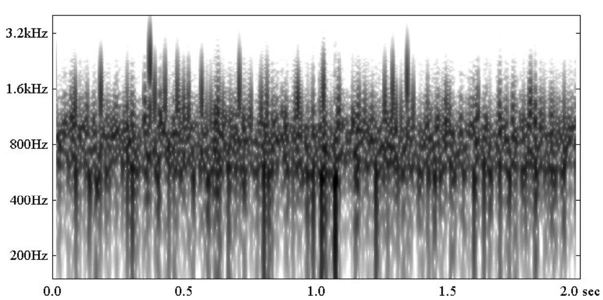
(a) Kurtosis = 1.



(b) Kurtosis = 2.

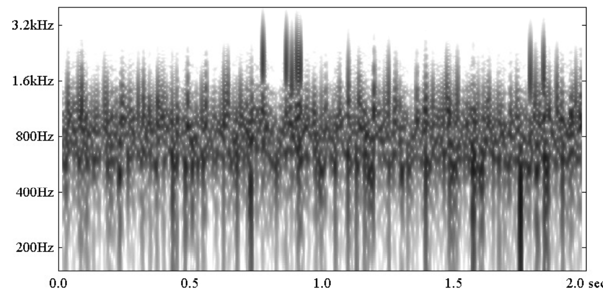


(c) Kurtosis = 4.

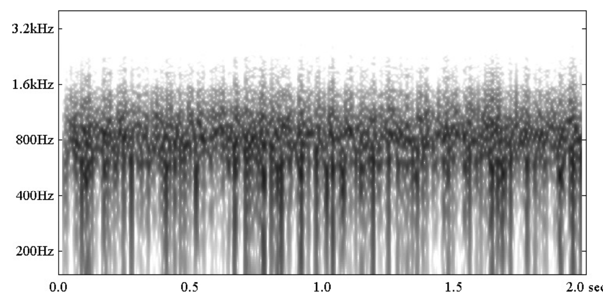


(c) Kurtosis = 8.

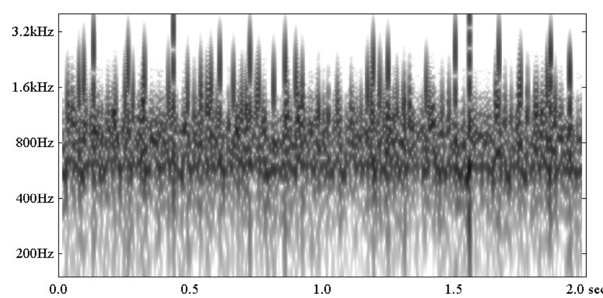
Figure 3: Spectrograms of the four consecutive sounds in Soundfile 1. The fixed parameters are: skewness = 0; $f_p = 800$ Hz; $c = 5/12$; $\Delta f = 600$ Hz.



(a) Skewness = -2.

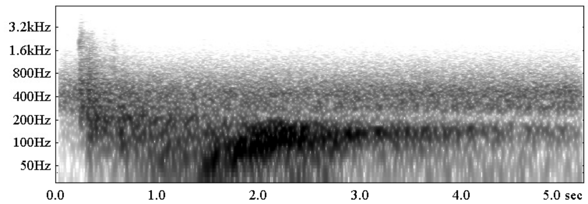


(b) Skewness = 0.

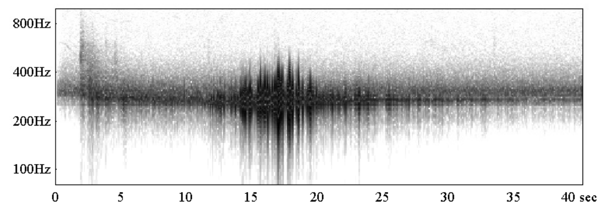


(c) Skewness = 2.

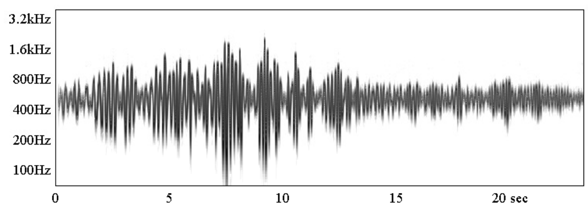
Figure 4: Spectrograms of the three consecutive sounds in Soundfile 2. The fixed parameters are: kurtosis = 12; $f_p = 800$; $c = 5/12$; $\Delta f = 600$.



(a) $c = 0$, $\Delta f = 0$ Hz, $f_p = 44.1$ kHz.

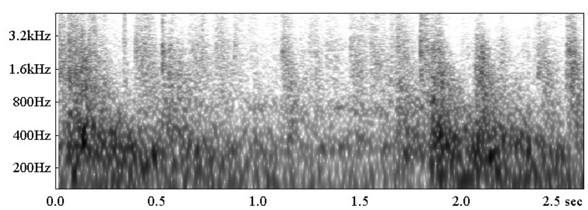


(b) $c = 1/12$, $\Delta f = 250$ Hz, $f_p = 5.5$ kHz.

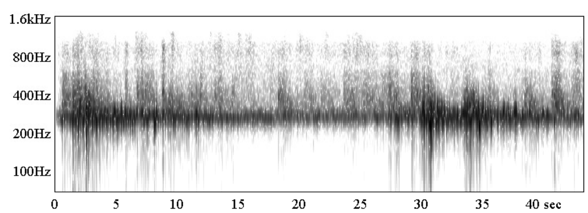


(c) $c = 3/12$, $\Delta f = 500$ Hz, $f_p = 2.75$ kHz.

Figure 5: Spectrograms of (a) pure audification, and (b and c) AugAudif of the example *Seismo1*.



(a) $c = 0$, $\Delta f = 0$ Hz, $f_p = 44.1$ kHz.



(b) $f_p = 2.75$kHz, $\Delta f = 250$Hz, $c = 1/12$.

Figure 6: Spectrograms of (a) pure audification, and (b) AugAudif of the example *Seismo2*.